

# Model Selection for Support Vector Machine Classification

Carl Gold

Computation and Neural Systems  
California Institute of Technology, 139-74  
Pasadena, CA 91125  
Email [carlg@caltech.edu](mailto:carlg@caltech.edu)

Peter Sollich

Department of Mathematics, King's College London  
Strand, London WC2R 2LS, U.K.  
Email [peter.sollich@kcl.ac.uk](mailto:peter.sollich@kcl.ac.uk)

May 3, 2002

We address the problem of model selection for Support Vector Machine (SVM) classification. For fixed functional form of the kernel, model selection amounts to tuning kernel parameters and the slack penalty coefficient  $C$ . We review a recently developed probabilistic framework for SVM classification. An extension to the case of SVMs with quadratic slack penalties is given and a simple approximation for the evidence is derived, which can be used as a criterion for model selection. We also derive the exact gradients of the evidence in terms of posterior averages and describe how they can be estimated numerically using Hybrid Monte Carlo techniques. Though computationally demanding, the resulting gradient ascent algorithm is a useful baseline tool for probabilistic SVM model selection, since it can locate maxima of the exact (unapproximated) evidence. We then perform extensive experiments on several benchmark data sets. The aim of these experiments is to compare the performance of probabilistic model selection criteria with alternatives based on estimates of the test error, namely the so-called “span estimate” and Wahba’s Generalized Approximate Cross-Validation (GACV) error. We find that all the “simple” model criteria (Laplace evidence approximations, and the Span and GACV error estimates) exhibit multiple local optima with respect to the hyperparameters. While some of these give performance that is competitive with results from other approaches in the literature, a significant fraction lead to rather higher test errors. The results for the evidence gradient ascent method show that also the exact evidence exhibits local optima, but these give test errors which are much less variable and also consistently lower than for the simpler model selection criteria.

